

Editor's column

Mark E. Segal

Chief, Computer and Information Sciences Research Group | Research Directorate | NSA

> The economics of computing continues to change in ways that allow larger computational problems to be solved at lower costs. Dramatic increases in commodity computing power, high-density disks that can store vast amounts of data, and very high-speed networks capable of moving information long distances very quickly are all making it possible to analyze the contents of massive data repositories and derive new insights from them. Many people refer to this state of technological evolution, coupled with the development of sophisticated new data-analysis algorithms, as the era of "Big Data." Big Data offers the promise of being able to detect trends in large data sets in ways that were not possible with older technologies.

Big Data capabilities can be applied to a wide variety of problems in many different domains. For example, in a Big Data world, it may be possible to provide better health care by detecting how diseases propagate in large populations. Big Data capabilities may also allow companies to spot new consumer trends in order to make products that people want to buy, and manufacture them in sufficient quantities so that everyone who wants the product can buy it when they want it.

In the scientific world, Big Data capabilities are making it possible to sift through vast quantities of data from sensors, such as weather satellites and particle accelerators, to increase our understanding of the physical world. Big Data capabilities can enhance national security by allowing our military to gain better situational awareness before and during a battle. Big Data capabilities may also be used to







analyze potential actions by a country or terrorist organization hostile to the United States and prevent those actions from taking place.

In this issue of The Next Wave (TNW), NSA researcher Paul Burkhardt provides an overview of Big Data, some of the key technologies behind it, and some of the key innovators in the field. One technological aspect of Big Data that is relevant to a wide variety of problems is the ability to analyze very large graphs. Burkhardt's second article discusses these "Big Graphs," showing how large-graph algorithms can be applied to several kinds of Big Data problems.

For the results of Big Data analysis to be useful to humans trying to solve difficult real-world problems, they must be put into a form that humans can understand and process. In the third article in this issue, NSA researchers Randall Rohrer, Celeste Paul, and Bohdan Nebesh explore this topic and discuss the connection between data visualization and analysis.

As Big Data analytics become more ubiquitous, concerns naturally arise about how data is collected, analyzed, and used. In particular, people whose data is stored in vast data repositories, regardless of who owns the repositories, are worried about potential privacy rights violations. Although privacy issues are not discussed in detail in this issue of TNW, an excellent overview of the relevant issues may be found in a report titled "Big Data and privacy: A technological perspective" authored by the President's Council of Advisors on Science and Technology and delivered to President Obama in May 2014 [1]. Another useful resource on this topic and

other topics related to Big Data is the article "Big Data and its technical challenges" by H. V. Jagadish et al. published in the July 2014 issue of *Communications of the ACM* [2].

According to a 2012 study by the International Data Corporation, there will be approximately 10²² bytes of data stored in all of the computers on Earth by 2015 [3]. To put that number in perspective, that's more than the estimated

7.5 x 10¹⁸ grains of sand on all of the beaches of the Earth [4], and almost as much as the estimated 10²² to 10²⁴ stars in the Universe [5, 6]. Let's harness the tools and algorithms currently being used to process Big Data to solve some of our planet's most critical problems. We hope you find this issue of *TNW* interesting, informative, and thought-provoking.

Mark E. Segal

Chief, Computer and Information Sciences Research Group Research Directorate, NSA

References

- [1] President's Council of Advisors on Science and Technology. "Big Data and privacy: A technological perspective." 2014 May 01. Available at: http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf.
- [2] Jagadish HV, Gehrke J, Labrinidis A, Papakonstantinou Y, Patel JM, Ramakrishnan R, Shahabi C. "Big Data and it's technical challenges." *Communications of the ACM*. 2014;57(7):86–94. doi: 10.1145/2611567.
- [3] Gantz J, Reinsel D. "The digital universe in 2020: Big Data, bigger digital shadows, and biggest growth in the Far East." International Data Corporation. 2012 Dec. Available at: http://idcdocserv.com/1414.
- [4] McAllister H. "Grains of sand on the world's beaches" [accessed 2014 Jun 16]. *The Journal of Modern Problem Solving*. Available at: http://www.hawaii.edu/suremath/jsand.html. (Problem was solved using SureMath software.)
- [5] Cain F. "How many stars are there in the Universe?" *Universe Today.* 3 Jun 2013. Available at: http://www.universetoday.com/102630/how-many-stars-are-there-in-the-universe/.
- [6] European Space Agency. "How many stars are there in the Universe?" [accessed 2014 Jun 16]. Available at: http://www.esa.int/Our_Activities/Space_Science/Herschel/How_many_stars_are_there_in_the_Universe.

The Next Wave is published to disseminate technical advancements and research activities in telecommunications and information technologies. Mentions of company names or commercial products do not imply endorsement by the US Government.

This publication is available online at http://www.nsa.gov/research/tnw/index.shtml. For more information, please contact us at TNW@tycho.ncsc.mil.